Marginal or conditional regression models for correlated non-normal data?

Muff, Stefanie ; Held, Leonhard ; Keller, Lukas F.

DOI: https://doi.org/10.1111/2041-210X.12623

Originally published at:
Muff, Stefanie; Held, Leonhard; Keller, Lukas F. (2016). Marginal or conditional regression models for correlated non-normal data? Methods in Ecology and Evolution, 7(12):1514-1524.
DOI: https://doi.org/10.1111/2041-210X.12623

Marginal or conditional regression models for correlated non-normal data?

Stefanie Muff[1,2⋆], Leonhard Held[2] and Lukas F. Keller[1]

[1] Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

[2] Department of Biostatistics, Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Hirschengraben 84, 8001 Zurich, Switzerland

⋆ Corresponding author. E-mail: stefanie.muff@uzh.ch, Phone: +41 44 634 48 16

1. Correlated data are ubiquitous in ecological and evolutionary research, and appropriate statistical analysis requires that these correlations are taken into account. For regressions with correlated, non-normal outcomes, two main approaches are used: conditional and marginal modelling. The former leads to generalized linear mixed models (GLMMs), while the latter are estimated using generalized estimating equations (GEEs), or marginalized multilevel regression models. Differences, advantages and drawbacks of conditional and marginal models have been discussed extensively in the statistical and applied literature, and there is some agreement that the choice of the model must depend on the question under study. Yet, there still appears to be a lot of confusion and disagreement over when to choose which model.

2. We start with a review of conditional and marginal models, and the differences in the interpretation of the resulting parameter estimates. We highlight that the two types of models propagate different linear relations between the covariates and the response. Moreover, while conditional models explicitly account for heterogeneity among clustered observations, marginal models yield averages over such heterogeneities and are therefore often interpreted as population-averaged models.

3. We point out theoretically and with an example that when modelling non-normal outcomes no unambiguous definition of a marginal model generally exists. Instead, marginal model parameters are marginal only with respect to unaccounted differences among clusters and thus depend on the fixed effects in the model. Therefore, marginal model parameters should not be loosely interpreted as population-averaged parameters. In addition, we explain

how marginal modelling is mathematically analogous to deliberately omitting covariates with explanatory power, and to deliberately introducing a Berkson measurement error into covariates. We also reiterate that marginal modelling is related to a well-known statistical phenomenon, the Simpson's paradox.

4. In most cases, therefore, we regard the conditional model as the more powerful choice to explain how covariates are associated with a non-normal response. Still, marginal models can be useful, given that the scientific question explicitly requires such a model formulation.

Keywords: Conditional model, generalized estimating equations, generalized linear mixed model, attenuation, Berkson measurement error, omitted covariates, Simpson's paradox

Introduction

Measurements and observations in ecology and evolution are often correlated, for example when repeated measurements or observations are recorded in longitudinal studies from the same individuals or populations, or when measurements are taken in temporal or spatial proximity (Zuur *et al.*, 2009; Fieberg *et al.*, 2010; Hamel *et al.*, 2012). Such correlations need to be accounted for in the analysis, as parameter estimates and their uncertainty can otherwise be biased. Because clustered and longitudinal data are ubiquitous in many research areas, their analysis has received tremendous attention in the past decades (Lindsey, 1999; Diggle *et al.*, 2002; Molenberghs & Verbeke, 2005; Fitzmaurice *et al.*, 2004).

A researcher assessing correlated data is usually faced with the decision whether it is more appropriate to use a *marginal* model, for instance by using a generalized estimating equations (GEEs) approach (Zeger & Liang, 1986; Liang & Zeger, 1986; Zeger *et al.*, 1988) or marginalized multilevel models (Heagerty, 1999; Heagerty &

Zeger, 2000; Schildcrout & Heagerty, 2007), or if it is better to formulate a *conditional* model, *i. e.* a generalized linear mixed model (GLMM). Conditional models include random effects to account for correlations within clusters, while marginal models require additional modelling steps to capture the dependencies. Historically, it was first possible to obtain robust estimates and to fit models for reasonably large data sets using GEEs, while GLMM inference only became feasible later. Hence, in practice, many researchers used marginal models because only they could be fit to correlated data. After it became viable to fit conditional models to complex and large data sets, a long discussion ensued in the statistical literature, whether to choose a conditional or a marginal model (Neuhaus *et al.*, 1991; Lindsey & Lambert, 1998; Heagerty & Zeger, 2000; Diggle *et al.*, 2002; Lee & Nelder, 2004). This discussion is still ongoing, also in ecology (Fieberg *et al.*, 2009; Koper & Manseau, 2009; Fieberg *et al.*, 2010; Akanda & Alpizar-Jara, 2014).

Whether a conditional or a marginal model is more appropriate for a specific research question has not been (and perhaps cannot be) answered in general. Lindsey & Lambert (1998) and Lee & Nelder (2004) discuss shortcomings of marginal modelling approaches, while text books like Zuur *et al.* (2009) or Hardin & Hilbe (2012) recommend them, and in particular GEEs, for correlated non-normal data. Marginal models fit by GEEs are hence popular in ecological applications (Keller *et al.*, 2002; Paradis & Claude, 2002; Bowerman *et al.*, 2010; Poncet *et al.*, 2010; Akanda & Alpizar-Jara, 2014).

For normally distributed response variables, *i. e.* in linear regression, the choice between a marginal and a conditional formulation is not particularly delicate, because the interpretation of conditional and marginal linear regression models turns out to be equivalent. On the other hand, the choice is relevant for non-normal data, as the interpretation of conditional and marginal regression models is usually different. A prominent view of conditional models is that they describe how the change of a covariate affects the response within a cluster, while holding the other covariates and all random effects constant. As clusters in ecology are often subjects (*e. g.*

individuals with repeated observations), conditional models are sometimes called *subject-specific* models. Here, however, we omit this naming, because it may imply an over-interpretation of the effects, particularly for covariates that do not vary within subjects (*e. g.* sex of an animal). On the other hand, marginal models are considered to be appropriate when inference on the population level is desired, irrespective of potential inter-cluster differences, and are therefore often denoted as *population-averaged* models.

The aims of this paper are twofold. First, we explain aspects and differences of conditional and marginal models which, we hope, will empower the users of these models. In particular, we discuss that the two models propagate different types of linear dependencies between the covariates and the response, and that the choice of a marginal vs. a conditional model must depend upon what the user considers to be a realistic model. Second, we scrutinize the interpretation of marginal models as population-averaged models and challenge the view that marginal models automatically answer reasonable marginal questions.

To this end, we first provide an introductory example, followed by an introduction to conditional and marginal models. We then briefly recapitulate known caveats of marginal models. We move on to point out that marginal models are related to seemingly disjoint topics. For example, marginal modelling is mathematically equivalent to deliberately omitting cluster-specific covariates or interaction terms (Neuhaus *et al.*, 1991). Moreover, marginal modelling is also analogous to deliberately introducing (or not accounting for) a particular type of measurement error in covariates, the so-called Berkson error (Berkson, 1950). We will only briefly touch on a third equivalence, that to the Simpson's paradox (Simpson, 1951), an equivalence that had been noted by Lindsey & Lambert (1998).

These equivalences highlight some of the conceptual difficulties of marginal models. They also point to another important aspect: the fact that a specific marginal model is only marginal with respect to unaccounted differences (heterogeneity) among clusters, *e. g.* with respect to covariates that are not included as fixed effects. Already Lindsey (1999) stated that "*All* statistical models are marginal with respect to any observed or unobserved covariates not used." Marginal parameter estimates must thus be interpreted accordingly. Our considerations will lead us to a similar conclusion as Lee & Nelder (2004), who "regard the conditional model as the fundamental, from which marginal predictions can be made."

Example: Mallard nest structures

This introductory example aims to illustrate differences between conditional and marginal model parameters for clustered data and to show how model parameters depend on the (cluster-specific) covariates included as fixed effects in the regression. We used a dataset of mallard nest structures and occupancy that was originally presented in Zicus *et al.* (2003, 2006) and then used by Fieberg *et al.* (2009) to illustrate the attenuation effect in marginal logistic regression. The scope of the original study was to investigate factors that explained mallard nest structure occupancy in Western Minnesota, using a binary response that indicated whether a nest structure was used (1) or not (0). Between 1997 and 1999, the same nest structures were inspected up to 4 times a year (periods 1–4), thus each nest structure i was observed at up to 12 time points j. The data therefore formed clusters of j observations per nest structure i. Covariates were observed at the level of the nest structures and included the amount of nesting cover, denoted as visual obstruction measurements (VOM), and the size of the open-water area in which the nesting structure was situated (size). Both VOM and size were assessed only once and are thus time-invariant within nest structure. In contrast to the analysis in Fieberg *et al.* (2009), we only included the 67 nests that were surrounded by an open-water area with a size between 0 and 25 km^2 and used the log-transformed, centered and standardized values of VOM and size (the reason for this will become apparent in the later sections). Fixed effects for period, year and their interaction were included as dummy variables. In a first comparison (model 1), we fitted a logistic regression using both a marginal and a conditional model. To account for correlations between repeated observations of the same nest structure, we used a GEE approach with an exchangeable correlation structure for the marginal model, and a nest-specific random intercept distributed as $\mathcal{N}(0, \sigma_\tau^2)$ for the GLMM. The models were fitted with the R software version 3.0.3 (R Core Team, 2013) using the functions `gee()` from the `gee` package (Caray *et al.*, 2015) and `glmer()` from the `lme4` package (Bates *et al.*, 2014). The code can be found in Appendix S1. As expected and already reported by Fieberg *et al.* (2009), the marginal parameter estimates (column GEE1 of Table 1) were are attenuated compared to their conditional counterparts (column GLMM1). This can also be seen in the fifth column of Table 1, where the ratios of the GEE1 to the GLMM1 estimates are given by the attenuation factor λ_1 for each parameter. The average of λ_1 was 0.860.

We then removed the nest structure-specific covariate `log(size)` for illustration purposes in a second model (model 2), and repeated the above analysis, resulting in the estimates GEE2 and GLMM2. The factor λ_2 denotes the attenuation of the GEE2 estimates relative to those obtained with GLMM2. Note that $\lambda_2 < \lambda_1$ for all coefficients with an average of 0.835 for λ_2, meaning that the marginal model parameters were slightly further attenuated due to the omission of the covariate `log(size)`. In contrast, the conditional model parameters from the two GLMM models were relatively similar. Importantly, the omission of the nest structure-specific covariate `log(size)` led to an increase in the among-nest variance estimates $\hat{\sigma}_\tau^2$ for model 2 (Table 1), which has a very intuitive interpretation: If less heterogeneity is explained by the fixed effects in the regression, more heterogeneity is captured by the nest structure-specific random intercept. Thanks to this feature, the conditional model estimates are quite robust with respect to the omission of the nest structure-specific, approximately Gaussian distributed covariate `log(size)`. On the other hand, the estimated working correlations $\hat{\rho}$ from the GEEs are almost the same for both models, with a small increase that is apparent only when allowing for three decimal places.

This example illustrates two main points to which we will return. First, the set of covariates included as fixed effects directly affects the estimates of the marginal model parameters, and thus the degree of attenuation compared to the conditional model increases when the cluster-specific covariate `log(size)` is excluded from the model. Second, both GEE1 and GEE2 are marginal models but because the unexplained heterogeneity between nest structures is larger in model 2 than in model 1, they are marginal with respect to a different residual component. Marginal model parameters should therefore be carefully interpreted in reference to the fixed effects that are (or are not) included in the regression model.

Conditional and marginal models

Before discussing the differences in parameter interpretation between marginal and conditional models in detail, we will first introduce the two types of models more formally. Let $\boldsymbol{x}_{ij}$ represent a (time-invariant) covariate vector for individual j in a cluster or a group i with $i = 1, \ldots, m$ and $j = 1, \ldots, n_i$. Alternatively, j is an index for repeated observations of subject i in a longitudinal study. Both cases typically result in a within-cluster correlation among the response variables y_{ij}, assumed here to be of exponential family form. One way to capture the correlation within clusters is to include a cluster-specific random effect vector $\boldsymbol{b}_i$ in the linear predictor

$$\eta_{ij} \,|\, \boldsymbol{b}_i \quad = \quad \beta_0 + \boldsymbol{x}_{ij}^\top \boldsymbol{\beta}_x + \boldsymbol{z}_{ij}^\top \boldsymbol{b}_i \ , \qquad \text{eqn 1}$$

which is linked to the mean $\mathsf{E}(y_{ij} \,|\, \boldsymbol{b}_i)$ via the inverse link function $h(\cdot)$ as $\mathsf{E}(y_{ij} \,|\, \boldsymbol{b}_i) = h(\eta_{ij} \,|\, \boldsymbol{b}_i)$. The components of the linear predictor are the intercept β_0, the fixed effects vector $\boldsymbol{\beta}_x$, the covariate vector $\boldsymbol{x}_{ij}$, and the cluster-specific random effects $\boldsymbol{b}_i$ that are linked to a covariate vector $\boldsymbol{z}_{ij}$. The random effects are assumed to be multivariate normal $\boldsymbol{b}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{D})$ with covariance matrix $\boldsymbol{D}$, and may incorporate group-specific, temporal or spatial dependencies. We explicitly condition on the random effects $\boldsymbol{b}_i$ in eqn 1 to stress that the model is *conditional*. The assumption is that, conditionally on $\boldsymbol{b}_i$ and the covariates, the y_{ij} are independent. To keep mathematics and notation simple, we will usually only include a Gaussian random intercept $\tau_i \sim \mathcal{N}(0, \sigma_\tau^2)$ and sometimes a random slope term for a single covariate with $b_i \sim \mathcal{N}(0, \sigma_b^2)$. Note that we use the term *conditional* here exclusively for a model that conditions on random effects, although the term can be used in a broader context also to condition on previous observations (Diggle *et al.*, 2002; Molenberghs & Verbeke, 2005).

In contrast, the corresponding so-called *(direct) marginal model* uses the same link function, but omits any cluster-specific random effects $\boldsymbol{b}_i$ in the model for the (transformed) mean. The linear predictor with $\mathsf{E}(y_{ij}) = h(\eta_{ij}^m)$ is then given as

$$\eta_{ij}^m = \beta_0^m + \boldsymbol{x}_{ij}^\top \boldsymbol{\beta}_x^m \ . \qquad \text{eqn 2}$$

The regression coefficients and the linear predictor of the *m*arginal model are labelled accordingly. In contrast to the conditional model, the y_{ij} are not assumed to be independent after accounting for the covariates. Further modelling steps to capture the dependencies of observations are thus required (see below), but the correlation structure is then essentially regarded as a nuisance.

Parameter estimation

The parameters of the mixed model as given in eqn 1 can be estimated via likelihood or Bayesian methods (Breslow & Clayton, 1993; Gelman *et al.*, 1995; Pinheiro & Bates, 2000; Diggle *et al.*, 2002; Wakefield, 2013). Details of estimating procedures are beyond the scope of this paper, but guidelines for ecologists and evolutionary biologists have, for instance, been summarized by Bolker *et al.* (2009).

Two main approaches to estimate the parameters of the marginal model (eqn 2) have been proposed. The most prominent is probably direct marginal modelling using GEEs (Zeger & Liang, 1986; Liang & Zeger, 1986; Zeger *et al.*, 1988), a quasi-likelihood approach (Wedderburn, 1974) that, in addition to the mean (eqn 2), requires the specification of a working correlation structure for the response. The estimating equations are solved numerically, however they do generally not correspond to any likelihood, hence the approach is termed quasi-likelihood. Another route to marginal model estimation involves the formulation of marginalized multilevel regression models that allow for likelihood-based marginal inference (Heagerty, 1999; Heagerty & Zeger, 2000; Schildcrout & Heagerty, 2007). Similarly to conditional models, marginalized multilevel regression models account for the dependence among measurements by introducing random effects, which allows for likelihood-based inference. In contrast to conditional models, however, the linear predictor is not conditioned on these random effects, so that the parameters still have a marginal interpretation.

Interpretation of the parameters

The main difference in the interpretation of conditional and marginal model parameters is the following: The marginal model assumes a linear relationship of the (transformed) mean with the covariates only (eqn 2), while the conditional model assumes a linear relationship of the (transformed) mean with the covariates *and* the random effects $\boldsymbol{b}_i$ (eqn 1). The latter capture unobserved characteristics of the clusters. More formally, the conditional parameter estimates for $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_x^\top)^\top$ of eqn 1 must be interpreted *conditionally* on cluster-specific values $\boldsymbol{b}_i$ that control for unknown heterogeneity and describe how a subject j in cluster i responds to a change in a covariate, while holding the other covariates and all random effects constant. This is similar in spirit to an experiment, where all conditions but the one of interest are fixed (Clayton & Hills, 1993, chapter 27). Adding random effects to the linear predictor of a conditional model automatically accounts for within-cluster similarity *and* between-cluster heterogeneity. In contrast, the marginal model parameters $\boldsymbol{\beta}^m = (\beta_0^m, \boldsymbol{\beta}_x^{m\top})^\top$ of eqn 2 are estimated *unconditionally* on any heterogeneity between clusters that is not captured by $\boldsymbol{x}_{ij}$. The parameters $\boldsymbol{\beta}^m$ therefore describe how subject j in cluster i responds to a change in a covariate in an *averaged* sense, which is why marginal models are often interpreted as population-averaged models, although this interpretation may be misleading. We come back to this important point later in the paper.

It is crucial to understand that the conditional and the marginal models are generally incompatible (Raudenbush, 2008): While a simple direct marginal model, as the one given by eqn 2, often leads to a complicated and unrealistic conditional relationship between the covariates and the response (Lindsey & Lambert, 1998; Raudenbush, 2008), a linear conditional relationship as in eqn 1 may lead to an intractable marginal model. For example, marginalization (*i. e.* integration) over the random effects of a logistic regression model with random intercept leads to

$$\mathsf{E}(y_{ij}) = \int_{-\infty}^{\infty} \frac{\exp(\beta_0 + \boldsymbol{x}_{ij}^\top \boldsymbol{\beta}_x + \tau_i)}{1 + \exp(\beta_0 + \boldsymbol{x}_{ij}^\top \boldsymbol{\beta}_x + \tau_i)} \frac{\exp(\tau_i^2/2\sigma_\tau^2)}{\sqrt{2\pi\sigma_\tau^2}} d\tau_i \ ,$$

which cannot be represented in a basic closed form. Simple linear dependencies of the response on the covariates, as given in eqn 1 and eqn 2, can generally not be correct at the same time, and therefore the estimated parameters may differ (Zeger *et al.*, 1988; Neuhaus *et al.*, 1991; Diggle *et al.*, 2002; Ritz & Spiegelman, 2004). The choice between a marginal and a conditional formulation should therefore be based on whether it appears more realistic that the marginal or the conditional transformed mean is linear in the covariates. Lindsey & Lambert (1998) have a clear opinion about this, which we essentially share, namely:

> The biological validity of such a [marginal] model will generally be questionable, although it will obviously have to be considered case by case. All models are wrong [...] but we are arguing that models directly constructed to have simple marginals will be more wrong than those beginning with a reasonably simple conditional relationship.

For similar reasons, Lee & Nelder (2004) propagate to deduce marginal predictions from conditional models, *e. g.* by integration over the random effects.

However, conditional and marginal models are not always incompatible. Models with identity link function, most prominently linear regression, are *conjugate* to themselves (Lee & Nelder, 2001). That is, marginalization over the normally distributed random effects results in a linear model with the same regression parameters, and the correlation matrix for the marginal model can be explicitly derived from the conditional model (Molenberghs & Verbeke, 2005). For instance, in a linear mixed model with residual variance σ_ϵ^2 and a random intercept $\tau_i \sim \mathcal{N}(0, \sigma_\tau^2)$, the intracluster-correlation for the response, given the observed covariates $\boldsymbol{x}_{ij}$, is $\sigma_\tau^2/(\sigma_\tau^2 + \sigma_\epsilon^2)$. Note, however, that from a knowledge of the marginal model parameters, it is not possible to deduce the exact structure of the conditional model, as different conditional models can lead to the same marginal model. Another exception where conditional and marginal models are not incompatible are log-linear models, such as Poisson regression, where all parameters except the intercept are the same for the marginal and conditional models (Zeger *et al.*, 1988; Neuhaus *et al.*, 1991), although this only holds when the respective conditional model includes a random intercept but no random coefficients (Grömping, 1996). Finally, when the variances of the random effects approach zero (*e. g.* $\sigma_\tau^2 \approx 0$ in the random intercept model), the two parameter sets converge for any link function (see Koper & Manseau, 2009).

Main advantages and disadvantages

Advantages and disadvantages of conditional and marginal models have been discussed for decades (Zeger & Liang, 1986; Heagerty & Zeger, 2000; Lindsey & Lambert, 1998; Lee & Nelder, 2004). Here we only repeat two main points raised in this discussion.

One reason that is often given as a key advantage of the marginal approach is that marginal model parameters are less demanding to fit and more robust against model misspecification than their conditional counterparts (Zeger & Liang, 1986; Heagerty & Zeger, 2000; Overall & Tonidandel, 2004; Zuur *et al.*, 2009). As an example, even if the within-cluster correlation structure in the GEE procedure is misspecified, the marginal regression estimates are still consistent(Zeger & Liang, 1986; Hardin & Hilbe, 2012). In addition, marginal modelling does either not require distributional assumptions about the random effects (GEEs), or is relatively robust to their misspecification (marginalized models, Heagerty & Zeger, 2000). As a consequence, one common criticism of GLMMs is that they rely on additional assumptions, such as the normality of random effects, and might be sensitive to their violation. Moreover, while GLMMs are applicable in a broader context than marginal models, for instance in the presence of complex spatial and/or temporal correlations (Pinheiro & Bates, 2000; Diggle *et al.*, 2002; Wakefield, 2013), fitting a GLMM can be a difficult and computationally demanding task and the user must select among various model fitting procedures, such as penalized quasilikelihood (PQL, Breslow & Clayton, 1993), Laplace approximations (Raudenbush *et al.*, 2000), adaptive Gauss-Hermite quadrature (GHQ, Pinheiro & Chao, 2006), or Bayesian approaches using Markov chain Monte Carlo (MCMC) sampling (Gamerman, 1997) or integrated nested Laplace approximations (INLA, Rue *et al.*, 2009).

A second important aspect of the conditional versus marginal modelling debate concerns model selection. Although likelihood-based estimation is possible in marginalized multilevel models (Heagerty & Zeger, 2000), the use of GEEs is still widespread. However, the lack of a likelihood functions imposes problems when it comes to model selection. In fact, marginal modelling with GEEs may lead to misleading results because certain model assumptions, such as the absence of interactions, are difficult or impossible to check (Lindsey & Lambert, 1998; Lee & Nelder, 2004). Although the quasi-likelihood information criterion (QIC) has been proposed for this purpose by Pan (2001), applications revealed that QIC rarely identifies the correct correlation structure and is thus not trustworthy (Koper & Manseau, 2009). On the other hand, marginalized multilevel models rely on likelihood methods and can thus undergo model selection procedures just like GLMMs. Although model choice can still be relatively challenging, valid information criteria for likelihood approaches exist, such as a modified Schwarz criterion (Pauler, 1998), the conditional AIC (Hodges & Sargent, 2001), a predictive cross-validation approach (Braun *et al.*, 2014), or a generalized R^2 (Nakagawa & Schielzeth, 2013; Johnson, 2014).

Analogies to marginal models

In this section we discuss that marginal modelling is mathematically equivalent to some seemingly unrelated phenomena. Already Lindsey & Lambert (1998) illustrated how marginal model parameters may have a misleading interpretation that is equivalent to that of Simpson's paradox (Simpson, 1951). The paradox describes a statistical phenomenon where a regression analysis of a full dataset leads to a trend that is the reverse of what is obtained when each cluster is regressed separately. The most prominent example of the paradox is probably the Berkeley gender bias, which has been described by Bickel *et al.* (1975). We will not further address Simpson's paradox here, but instead concentrate on two others analogies to marginal modelling: that to omitted covariates or interaction terms, and that to (additive or multiplicative) Berkson measurement error.

Analogy to omitted covariates

It is well-known that estimates of regression models can change if important covariates are omitted (Lee, 1982; Gail *et al.*, 1984; Lee & Nelder, 2004), a phenomenon that is also known as non-collapsibility (Greenland *et al.*, 1999). Starting from the fact that conditional and marginal models for correlated data differ in how unobserved heterogeneity is modelled, we will argue that the use of a marginal model may be conceptually equivalent to deliberately omitting (observed or unobserved) covariates or interactions with covariates. This interpretation will then be used to illustrate that each set of covariates defines a model with different margins, thus there is no unambiguous definition of a marginal model.

Assume that a conditional model with a random intercept describes our data well. The random intercepts $\tau_i \sim \mathcal{N}(0, \sigma_\tau^2)$ can then be interpreted as realizations of a cluster-specific (unobserved) covariate, a view that was formalized by Neuhaus *et al.* (1991). To this end, substitute $\tau_i = \sigma_\tau z_i$ with standardized, cluster-specific covariate

$z_i \sim \mathcal{N}(0,1)$. The model can then be rewritten as

$$\eta_{ij} \,|\, z_i = \beta_0 + \boldsymbol{x}_{ij}^\top \boldsymbol{\beta}_x + \sigma_\tau z_i \ , \qquad \text{eqn 3}$$

where the standard deviation σ_τ may be interpreted as a regression coefficient for z_i. The difference in the linear predictors of the conditional and marginal models is then the inclusion or exclusion of τ_i, or, with this new interpretation, the inclusion or exclusion of the covariate z_i. Conceptually, omitting z_i from the linear predictor is equivalent to fitting a marginal model, irrespective of whether the values of z_i are known or unknown. In practice, however, if z_i is known, it would typically be directly included in the marginal model as

$$\eta_{ij}^m = \beta_0^m + \boldsymbol{x}_{ij}^\top \boldsymbol{\beta}_x^m + \beta_z^m z_i \ , \qquad \text{eqn 4}$$

and σ_τ in eqn 3 then corresponds to β_z^m. Assuming that no additional covariates are missing, conditional and marginal model equations are then identical, because no residual heterogeneity discriminates the models. It is thus not surprising that Gail *et al.* (1984) and Neuhaus *et al.* (1991) deduced qualitatively the same attenuation factors for the parameters of a binary regression model in two distinct contexts, once from the omitted covariate perspective, and once for omitted random effects.

Assume there exists another cluster-specific covariate w_i, which only partially explains the between-cluster heterogeneity of eqn 3, *i. e.*

$$z_i = w_i + u_i \ ,$$

with u_i independent of w_i, representing the residual between-cluster heterogeneity not explained by w_i, *i. e.* $w_i \sim \mathcal{N}(0, \sigma_w^2)$, $u_i \sim \mathcal{N}(0, \sigma_u^2)$ and $\sigma_w^2 + \sigma_u^2 = 1$. The model

$$\eta_{ij}^m = \beta_0^m + \boldsymbol{x}_{ij}^\top \boldsymbol{\beta}_x^m + \beta_z^m w_i$$

is then marginal with respect to the between-cluster heterogeneity carried by u_i, and a random term $\tau_i' \sim \mathcal{N}(0, \sigma_\tau^2 \sigma_u^2)$ in the respective conditional model

$$\eta_{ij} \,|\, \tau_i' = \beta_0 + \boldsymbol{x}_{ij}^\top \boldsymbol{\beta}_x + \beta_z w_i + \tau_i'$$

could capture it. Note that conditional and marginal models converge as the variance $\sigma_w^2 \to 1$, *i. e.* as the unexplained heterogeneity σ_u^2 approaches zero.

A similar reformulation as above can be used to show that the omission of a random coefficient with non-zero variance can be interpreted as a missing interaction term. Replacing the random slope $b_i \sim \mathcal{N}(0, \sigma_b^2)$ in the following equation that includes a single covariate x_{ij}

$$\eta_{ij} \,|\, b_i = \beta_0 + \beta_x x_{ij} + \tau_i + b_i x_{ij} \qquad \text{eqn 5}$$

by $b_i = \sigma_b z_i$ with standard normal z_i, leads to

$$\eta_{ij} \,|\, z_i = \beta_0 + \beta_x x_{ij} + \tau_i + \sigma_b z_i x_{ij} \ .$$

As above, the standard deviation σ_b can be interpreted as the coefficient for the interaction term $z_i x_{ij}$, thus omitting b_i is equivalent to omitting the interaction between z_i and x_{ij}.

These considerations illustrate two things. First, there is a straightforward analogy between marginal regression models and the omission of covariates or interaction terms, no matter if the respective variables were observed or not. And second, the distinction between conditional and marginal models is less clear than often suggested. In the random intercept model, for instance, the variance of τ_i strongly depends on the covariates included as fixed effects. All covariates that are constant or correlated within clusters potentially explain part of the between-cluster heterogeneity. The more such covariates are included, the smaller σ_τ^2 potentially is. Importantly, in cases where marginal model parameters are attenuated with respect to their conditional counterparts, as in logistic regression (Zeger *et al.*, 1988;

Neuhaus *et al.*, 1991), the degree of attenuation directly depends on σ_τ^2, and thus on the set of covariates. This dependence of the regression parameters on the set of covariates may be problematic if the aim is to assess how covariates are associated with the response, but many practitioners may not be aware of it. An illustration of this effect is evident in the mallard nest structure example, where the attenuation of the regression parameters increased when a relevant cluster-specific covariate was excluded from the marginal logistic regression model. The user of marginal regression models must thus appreciate that the respective parameters answer only one particular marginal question with very specific margins. Therefore, the interpretation of marginal model parameters as those of a population-averaged model is not necessarily meaningful because the model "averages" only over the heterogeneity that is not explained by the covariates. In practice, additional covariates are often collected with exactly the purpose to explain as much response variability as possible, hoping to obtain better effect estimates for the covariate(s) of interest. A model is then typically specified depending on the data that are available; the more covariates are observed, the more are included as fixed effects. From this perspective it appears unreasonable to ignore additional "unobserved covariates" by not including random effects in the model.

This discussed analogy between covariates and Gaussian random effects obviously only holds as long as the distribution of the respective covariate is also Gaussian. Note that this was the reason why we log-transformed the continuous variables in the mallard example (Figure 1). An important difference between a known, cluster-specific covariate and a random effect is that the former requires no additional modelling assumptions, while random effects do. In fact, the aspect of distributional assumptions is an important and justified criticism of conditional modelling (Neuhaus *et al.*, 1992; Heagerty & Zeger, 2000). Still, whether robustness considerations alone legitimate the use of models for the mean that do not account for between-cluster heterogeneity seems at least questionable.

Analogy to Berkson measurement error

As mentioned above and illustrated with the mallard example, marginal model parameters $\boldsymbol{\beta}^m$ may be attenuated with respect to the conditional parameters $\boldsymbol{\beta}$. Such attenuation effects are also known from measurement error theory (Stefanski, 1985; Fuller, 1987). Already Zeger *et al.* (1988) pointed out that there might be a relation, and Diggle *et al.* (2002) stated in the context of marginal modelling that the phenomenon of attenuation is well-known in the "related errors-in-variable regression problem". As we will show here, the relation of marginal modelling to measurement error theory only holds when the error in a covariate follows a specific type, namely the Berkson measurement error (Berkson, 1950). Under the additive and multiplicative Berkson error structure, a true covariate x is related to an error-prone observation w via $x = w + u$ or $x = w \cdot v$ with additive and multiplicative error terms $u \sim \mathcal{N}(0, \sigma_u^2)$ and $v \sim \mathcal{N}(1, \sigma_v^2)$, both assumed independent of w. Note that this error is different from the so-called classical error type, where $w = x + u$ and $w = x \cdot v$, respectively. Typical settings where Berkson-type errors occur in biological studies are the assignment of averages of exposures to individuals (plants, animals) growing or living in the vicinity of a measurement station, or in experimental setups. A prominent example is the application of fixed doses of herbicides in bioassay experiments (Rudemo *et al.*, 1989). In the following subsections we will illustrate how moving from a conditional to a marginal model by omitting a random intercept or a random slope is mathematically analogous to introducing an additive or a multiplicative Berkson error into a covariate.

Random intercept model

Let us assume that a random intercept $\tau_i \sim \mathcal{N}(0, \sigma_\tau^2)$ in the linear predictor of eqn 1 captures the heterogeneity between clusters. As noted by Rudemo *et al.* (1989) and Carroll *et al.* (2006, p.189), omitting such a random intercept is equivalent to introducing homoscedastic, additive Berkson error into a covariate. To see this, we use the model given in eqn 1 with a single covariate x_{ij} and a random intercept τ_i, and reformulate it to

$$\eta_{ij} \,|\, \tau_i = \beta_0 + \beta_x(x_{ij} + \tau_i/\beta_x) \ .$$

Additional covariates $\boldsymbol{z}_{ij}$ could be added to the above equation, but the random effect is attributed here only to the (freely chosen) covariate x_{ij}. This model can be interpreted as a fixed-effects generalized linear model (GLM) with true covariate $x^\star_{ij} = (x_{ij} + \tau_i/\beta_x)$. On the other hand, the linear predictor of the corresponding marginal model $\eta^m_{ij} = \beta^m_0 + \beta^m_x x_{ij}$ does not contain the random effect τ_i, and therefore includes x_{ij} as covariate instead of $x^\star_{ij}$. Given that we can rewrite $x^\star_{ij} = x_{ij} + u_{ij}$ with $u_{ij} = \tau_i/\beta_x$, distributed as

$$u_{ij} \sim \mathcal{N}(0, \sigma^2_\tau/\beta^2_x) \ , \qquad \text{eqn 6}$$

the inclusion of x_{ij} instead of $x^\star_{ij}$ in the linear predictor is equivalent to an additive Gaussian Berkson error. Note that the error terms are constant within clusters i, as $u_{ij} = \tau_i/\beta_x$ for all individuals j in i.

Thanks to the above analogy it is straightforward to transfer results from Berkson error theory to the marginal modelling context, and vice versa. In fact several analogous results have been reported in parallel from those two perspectives. Two prominent examples are given below.

Binomial regression with probit link: In probit regression we have a binary outcome $y_{ij} \,|\, \tau_i \sim \text{Bernoulli}(p_{ij})$, $p_{ij} \,|\, \tau_i = \Phi(\beta_0 + \boldsymbol{x}^\top_{ij}\boldsymbol{\beta}_x + \tau_i)$, with cumulative standard normal distribution function Φ and $\tau_i \sim \mathcal{N}(0, \sigma^2_\tau)$. Zeger *et al.* (1988) and Heagerty & Zeger (2000) have shown that the parameters of the marginal and conditional models are related as $\boldsymbol{\beta}^m = \lambda\boldsymbol{\beta}$ with attenuation factor $\lambda = (1 + \sigma^2_\tau)^{-1/2}$. On the other hand, Burr (1988) and Tosteson *et al.* (1989) derived attenuation factors for the parameters in probit regression when the (single) covariate is subject to Berkson error. Denote the error-free covariate as $x^\star_{ij}$, while x_{ij} is the observed covariate subject to Berkson error with $x^\star_{ij} = x_{ij} + u_{ij}$ and error variance $u_{ij} \sim \mathcal{N}(0, \sigma^2_u)$. Then

$p_{ij} = \Phi(\beta_0 + \beta_x x_{ij}^\star)$ is the correct model with parameters β_0 and β_x, but if $x_{ij}^\star$ is replaced by its error-prone proxy x_{ij}, both parameters are attenuated by a factor $\lambda' = (1 + \beta_x^2 \sigma_u^2)^{-1/2}$ (Burr, 1988; Tosteson *et al.*, 1989). Using eqn 6 we know that $\sigma_\tau^2 = \beta_x^2 \sigma_u^2$, and therefore $\lambda = \lambda'$.

Binomial regression with logit link: Using the approximate relationship between probit and logistic regression (Johnson & Kotz, 1970), Zeger *et al.* (1988) have shown that the relation between the conditional and the marginal model parameters is given as $\boldsymbol{\beta}^m \approx (1 + c^2\sigma_\tau^2)^{-1/2}\boldsymbol{\beta}$ with $c = 16\sqrt{3}/(15\pi) \approx 0.588$. In analogy, Reeves *et al.* (1998) have shown that in the presence of Berkson error with $u_i \sim \mathcal{N}(0, \sigma_u^2)$, the estimated parameters in logistic regression are attenuated by the factor $(1 + c^2\beta_x^2\sigma_u^2)^{-1/2}$. Again, these results are equivalent due to eqn 6.

Random slope model

A similar equivalence as above, but for multiplicative Berkson error, can be deduced when the true conditional model includes a random slope. The analogy was briefly mentioned by Rudemo *et al.* (1989). For simplicity, we use the model as given in eqn 5, however ignoring the random intercept τ_i for the moment. This model can be reformulated as

$$\eta_{ij} \,|\, b_i = \beta_0 + \beta_x \left[x_{ij} \cdot \left(1 + \frac{b_i}{\beta_x}\right)\right] = \beta_0 + \beta_x x_{ij}^{\star\star} \ ,$$

where $x_{ij}^{\star\star} = x_{ij} \cdot (1 + b_i/\beta_x)$ can be interpreted as the true covariate of a GLM. Since the linear predictor of the respective marginal model does not include b_i, the replacement of $x_{ij}^{\star\star}$ by x_{ij} corresponds to replacing the conditional by the marginal linear predictor. On the other hand, due to $x_{ij}^{\star\star} = x_{ij} \cdot u_{ij}$ with $u_{ij} = 1 + b_i/\beta_x$ distributed as $u_{ij} \sim \mathcal{N}(1, \sigma_b^2/\beta_x^2)$, replacing $x_{ij}^{\star\star}$ by x_{ij} also corresponds to introducing multiplicative, homoscedastic Berkson error into the covariate. Again, u_{ij} is constant for all individuals within a cluster. Most realistic conditional models will have both a random intercept *and* a random slope, as in eqn 5. The marginal counterpart then omits both terms simultaneously, leading to a mixture of additive and multiplicative Berkson error.

Comparisons between marginal and conditional model parameters in the presence of random slope coefficients are rare in the literature. An interesting exception is Grömping (1996), who gave a general expression for the relation of conditional and marginal parameters in the case of log-linear mean models, *e. g.* Poisson regression. In contrast to the case of log-linear models with random intercepts, Grömping (1996) showed that the omission of random coefficients with non-zero variances may lead to attenuation *or* reverse attenuation effects, and that non-intercept parameters may then be affected as well.

The mallard nest structure example revisited

We now briefly revisit the introductory example and enrich it with theoretical insight from the previous sections. The logistic model to explain mallard nest occupancy included as fixed effects the two continuous covariates `log(VOM)` and `log(size)`, as well as period, year and their interaction as dummy variables in the $\boldsymbol{x}_{ij}$ matrix. The marginal linear predictor is thus given by $\eta^m_{ij} = \log(p_{ij}/(1-p_{ij}))$ with $y_{ij} \sim$ Bernoulli(p_{ij}), and can be written as

$$\eta^m_{ij} = \beta^m_0 + \beta^m_1 \texttt{log(VOM)}_{ij} + \beta^m_2 \texttt{log(size)}_{ij} + \boldsymbol{x}_{ij}^\top \boldsymbol{\beta}^m_3 \ ,$$

where j denotes an observation of nest structure i, and $\boldsymbol{\beta}^m = (\beta^m_0, \beta^m_1, \beta^m_2, \boldsymbol{\beta}^{m\top}_3)^\top$ is the vector of marginal regression coefficients. The corresponding conditional regression model adds a nest structure-specific random intercept

$$\eta_{ij} \,|\, \tau_i = \beta_0 + \beta_1 \texttt{log(VOM)}_{ij} + \beta_2 \texttt{log(size)}_{ij} + \boldsymbol{x}_{ij}^\top \boldsymbol{\beta}_3 + \tau_i \ ,$$

with $\tau_i \sim \mathcal{N}(0, \sigma_\tau^2)$. For this model, which we termed model 1, we found an average attenuation factor between marginal and conditional parameter estimates of $\lambda_1 = 0.860$. Using the estimated variance $\hat{\sigma}_\tau^2 = 1.16$, the constant $c = 0.588$, and the formula of Zeger *et al.* (1988), we obtain a theoretical prediction of $\hat{\lambda}_1 = \boldsymbol{\beta}^m/\boldsymbol{\beta} \approx (1 + c^2\sigma_\tau^2)^{-1/2} = 0.845$. For model 2, where we removed the nest structure-specific and approximately Gaussian distributed covariate `log(size)` (Figure 1), an average attenuation factor of $\lambda_2 = 0.835$ was observed, while plugging $\hat{\sigma}_\tau^2 = 1.50$ from GLMM2 into the above formula yields $\hat{\lambda}_2 = 0.812$. The fact that $\hat{\lambda}_2 < \hat{\lambda}_1$ reiterates that omitting cluster-specific covariates increases the observed parameter attenuation in marginal models.

Let us examine these results also from the omitted covariates perspective. In eqn 3 and eqn 4 we have seen how σ_τ^2 absorbs between-cluster heterogeneity when a Gaussian, cluster-specific covariate (that is independent of the other explanatory variables) is omitted. Here, the variance estimate $\hat{\sigma}_\tau^2$ increases by 0.34 when switching from model 1 to model 2, which is comparable to the squared coefficient of the standardized omitted covariate $\hat{\beta}_2^2 = 0.46$.

Discussion

The fact that conditional and marginal regression models of non-normal data differ in their interpretation and may yield different parameter estimates has been extensively discussed in the literature. Still, differences as those among the parameter estimates in logistic regression keep being reported in ecological and medical statistics (Young *et al.*, 2007; Koper & Manseau, 2009; Gardiner *et al.*, 2009), and there appears yet to be no consensus about when or if a marginal model should be preferred to a conditional model (Fieberg *et al.*, 2009; Hubbard *et al.*, 2010).

Here, we tried to broaden the view on the interpretation of conditional and marginal models. In particular, we reviewed how three apparently unrelated topics are connected to marginal modelling: the omission of cluster-specific covariates or interactions, Berkson measurement error, and Simpson's paradox. While the analogy to the latter had been discussed previously (Lindsey & Lambert, 1998), and the equivalence of marginal models to omitted covariates has been mentioned by Neuhaus *et al.* (1991), the Berkson measurement error analogy was only pointed out in the context of dose-response curve models by Rudemo *et al.* (1989), and tangentially noted by Carroll *et al.* (2006). Moreover, an explicit relationship has, to our knowledge, never been established, so that several equivalent results were derived independently. Given these analogies, it is not a coincidence that the same types of models that are robust to marginalization (*e. g.* the linear regression model) are also robust to the omission of covariates and to Berkson measurement error. These alternative viewpoints substantiate concerns about marginal modelling of non-normal data that have been raised previously (Lindsey & Lambert, 1998; Lee & Nelder, 2004; Subramanian & O'Malley, 2010). In light of the analogies we discussed, one might want to scrutinize model formulations that seem reasonable otherwise. For example, it may appear sensible to the analyst to choose a marginal instead of a conditional model, for instance due to some robustness considerations. However, few analysts would deliberately introduce (or ignore) a Berkson measurement error, or omit a relevant covariate if it was possible to account for it.

The omitted-covariate view is useful to illustrate that the marginal model is not uniquely specified. Instead, what is often termed to be *the* marginal model is in fact one of many marginal models. For instance, when the conditional model includes only a cluster-specific random intercept, the between-cluster heterogeneity that is captured by the variance σ_τ^2 decreases as more covariates that are correlated or constant within clusters are included in the regression. In the extreme, σ_τ^2 converges to zero and the respective marginal and conditional models are identical. Each marginal model with a selected set of covariates thus reflects a population average only with respect to the population characteristics that are not represented as covariates in the model. In applications, the choice of covariates that are monitored during data collection often depends on practical considerations or technical limitations. Consequently, marginalization is usually over a very specific (and somewhat arbitrary) component of the response's between-cluster variability, which does not necessarily result in a relevant marginal interpretation, but in some sort of hybrid model that explicitly accounts for some (the observed) but not for another (the unobserved) part of the between-cluster heterogeneity. The interpretation of the respective marginal model parameters is thus difficult, in particular because effect size estimates of marginal models may depend on the heterogeneity among clusters that is not already absorbed by the linear predictor. We regard this as an undesirable property of marginal models, but many users are probably unaware of this aspect. A potential conclusion might be that a "true" marginal answer can only be obtained if all cluster-specific covariates (or those with a cluster-specific component) except the one of interest are excluded from the model. However, it seems unnatural (and most researchers might not be willing) to omit known covariates if they have explanatory power for the model.

Of course, we do not generally deny the usefulness of marginal regression models, because it may well be that a scientific question requires a direct marginal model formulation. If any of the questions "Is it realistic to propagate a linear relationship between the observed predictors and the transformed mean, but not for additional unobserved covariates or latent characteristics of the clusters?" or "Is it expected that all between-cluster variability has been captured by the covariates?" can be answered affirmatively, a marginal model may be useful. Ideally, such assumptions should be substantiated or checked, *e. g.* by examining residual correlations. In any case, where possible the choice between a marginal and a conditional formulation should not be based on the availability of a convenient software solution, *e. g.* R-libraries.

Sometimes marginal models are propagated as the models of choice when the aim is prediction rather than explanation. However, we believe that a realistic explanatory model is an important starting point for realistic predictions. For the same reason, Lee & Nelder (2004) proposed to deduce marginal predictions from conditional models, *i. e.* by integrating out the heterogeneities among clusters. No matter whether the explanatory model is marginal or conditional, predictions for populations other than the study sample are only valid if the latter is a random sample for the predicted population, at least with respect to the heterogeneities that could not be captured by the fixed effects. A violation of this condition can lead to invalid, that is, biased predictions under both conditional and marginal models.

An important limitation of the considerations in this article is the restriction to time-invariant covariates. Note that model interpretation and estimating procedures are very different when covariates or covariate effects are allowed to be time-dependent in longitudinal data analysis (Pepe & Anderson, 1994; Diggle *et al.*, 2002; Schildcrout & Heagerty, 2005). Such aspects are beyond the scope of this paper. Moreover, our considerations were limited to independent Gaussian random effects, and therefore also to Gaussian Berkson measurement error and Gaussian omitted covariates and interaction terms. The normality assumption is commonly used in practice, albeit its violation can lead to biased parameter estimates (Neuhaus *et al.*, 1992; Heagerty & Zeger, 2000). Of course, random effects may sometimes be non-Gaussian, or feature a temporal, spatial or phylogenetic correlation (Ives & Helmus, 2011; Kaldhusal *et al.*, 2015; Hadfield, 2015). The above analogies are then still valid, but the distribution and the dependencies must be passed on to the respective Berkson error terms or to the respective omitted covariate. Nevertheless, the specification of distributional assumptions in conditional models is clearly critical. The robustness of the marginal approach against such misspecification is sometimes used as an argument to prefer a marginal model formulation. However, we argue that if the conditional model represents a more realistic relationship between the covariates and the response, which may indeed often be the case, there is little benefit from robustly estimated, but biased parameters.

A practitioner might now wonder whether it is advisable to include random effects of various kinds by default in order to not miss any heterogeneity among clusters. However, this is not the message of this article, and we do not at all advocate the promiscuous inclusion of random effects. As in any data analysis procedure, the user must specify a realistic model including covariates and random effects from a priori knowledge of causal relations and, if necessary, use a suitable model validation approach to improve the model.

In summary, we have discussed how correlated non-normal data can be analyzed using conditional or marginal models, but that the parameter interpretation can be quite different under these two models. We highlighted that moving from a conditional to a marginal model is mathematically analogous to deliberately introducing an additive or a multiplicative Berkson measurement error, or to deliberately omitting cluster-specific covariates. Looking at marginal models through these analogies makes it clear that while marginal models may be justified under certain circumstances, conditional models will most often be the preferred choice. If marginal conclusions are required, they can be obtained by integrating over the conditional models.

Acknowledgements

The authors thank Prof. Torsten Hothorn, Dr. Beate Sick and Benedikt Gehr for useful comments on the manuscript, and two anonymous reviewers who helped to significantly improve the paper. This work would not have been possible without generous funding from the Faculty of Science of the University of Zurich.

Data Accessibility

The R script to reproduce the mallard nest structures example is uploaded as online supporting information (Appendix S1). The respective data set was taken from Appendix S3 of Fieberg *et al.* (2009), which can be downloaded from
http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2664.2009.01692.x/suppinfo

References

Akanda, M.A.S. & Alpizar-Jara, R. (2014) Estimation of capture probabilities using generalized estimating equations and mixed effects approache. *Ecology and Evolution*, **4**, 1158–1165.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2014) *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-7.

Berkson, J. (1950) Are there two regressions? *Journal of the American Statistical Association*, **45**, 164–180.

Bickel, P.J., Hammel, E.A. & O'Connell, J.W. (1975) Sex bias in graduate admissions: Data from Berkeley. *Science*, **4175**, 398–404.

Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H. & White, J.S.S. (2009) Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology and Evolution*, **24**, 127–135.

Bowerman, J., Johnson, P.T.J. & Bowerman, T. (2010) Sublethal predators and their injured prey: linking aquatic predators and severe limb abnormalities in amphibians. *Ecology*, **91**, 242–251.

Braun, J., Sabanés Bové, D. & Held, L. (2014) Choice of generalized linear mixed models using predictive crossvalidation. *Computational Statistics & Data Analysis*, **75**, 190–202.

Breslow, N.E. & Clayton, D.G. (1993) Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**, 9–25.

Burr, D. (1988) On errors-in-variables in binary regression—Berkson case. *Journal of the American Statistical Association*, **83**, 739–743.

Caray, V.J., Lumley, T. & Ripley, B. (2015) *Package 'gee'*. R package version 4.13-19.

Carroll, R., Ruppert, D., Stefanski, L. & Crainiceanu, C. (2006) *Measurement Error in Nonlinear Models: A Modern Perspective.* Chapman & Hall, Boca Raton, 2 edition.

Clayton, D. & Hills, M. (1993) *Statistical Models in Epidemiology.* Oxford University Press, Oxford.

Diggle, P.J., Heagerty, P.J., Liang, K.Y. & Zeger, S.L. (2002) *Analysis of Longitudinal Data.* Oxford Science Publications. Clarendon Press, Oxford, 2 edition.

Fieberg, J., Matthiopoulos, J., Hebblewhite, M., Boyce, M.S. & Frair, J.L. (2010) Correlation and studies of habitat selection: problem, red herring or opportuniy? *Phil Trans R Soc B*, **365**, 2233–2244.

Fieberg, J., Rieger, R.H., Zicus, M.C. & Schildcrout, J.S. (2009) Regression modelling of correlated data in ecology: subject-specific and population averaged response patterns. *Journal of Applied Ecology*, **46**, 1018–1025.

Fitzmaurice, G.M., Laird, N.M. & Ware, J.H. (2004) *Applied Longitudinal Analysis.* Wiley Series in Probability and Statistics. John Wiley & Sons: New Jersey.

Fuller, W.A. (1987) *Measurement Error Models.* John Wiley & Sons, New York.

Gail, M.H., Wieand, S. & Piantadosi, S. (1984) Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika*, **71**, 431–444.

Gamerman, D. (1997) Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing*, **7**, 57–68.

Gardiner, J.C., Luo, Z. & Roman, L.A. (2009) Fixed effects, random effects and GEE: what are the differences? *Statistics in Medicine*, **28**, 221–239.

Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B. (1995) *Bayesian Data Analysis.* Texts in Statistical Science. London: Champan & Hall.

Greenland, S., Robins, J.M. & Pearl, J. (1999) Confounding and collapsibility in causal inference. *Statistical Science*, **14**, 29–46.

Grömping, U. (1996) A note on fitting a marginal model to mixed effects log-linear regression data via GEE. *Biometrics*, **52**, 280–285.

Hadfield, J.D. (2015) Increasing the efficiency of MCMC for hierarchical phylogenetic models of categorical traits using reduced mixed models. *Methods in Ecology and Evolution*, **6**, 706–714.

Hamel, S., Yoccoz, N.G. & Gaillard, J.M. (2012) Statistical evaluation of parameters estimating autocorrelation and individual heterogeneity in longitudinal studies. *Methods in Ecology and Evolution*, **3**, 731–742.

Hardin, J.W. & Hilbe, J.M. (2012) *Generalized Estimating Equations*. Chapman & Hall/CRC, Boca Raton, 2 edition.

Heagerty, P. (1999) Marginally specified logistic-normal model for longitudinal binary data. *Biometrics*, **55**, 688–698.

Heagerty, P. & Zeger, S. (2000) Marginalized multilevel models and likelihood inference (with comments and a rejoinder by the authors). *Statistical Science*, **15**, 1–26.

Hodges, J.S. & Sargent, D.J. (2001) Counting degrees of freedom in hierarchical and other richly-parameterised models. *Biometrika*, **88**, 367–379.

Hubbard, A.E., Ahern, J., Fleischer, N.L., Van der Laan, M., Lippman, S.A., Jewell, N., Bruckner, T. & Satariano, W.A. (2010) To GEE or not to GEE: comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. *Epidemiology*, **21**, 467–474.

Ives, A.R. & Helmus, M.R. (2011) Generalized linear mixed models for phylogenetic analyses of community structure. *Ecological Monographs*, **81**, 511–525.

Johnson, N.L. & Kotz, S. (1970) *Distribution in Statistics, Continuous Univariate Distributions*. Houghto-Mifflin, Boston, 2 edition.

Johnson, P.C.D. (2014) Extension of Nakagawa & Schielzeth's R^2_{GLMM} to random slopes models. *Methods in Ecology and Evolution*, **5**, 944–946.

Kaldhusal, A., Brandl, R., Müller, J., Möst, L. & Hothorn, T. (2015) Spatio-phylogenetic multispecies distribution models. *Methods in Ecology and Evolution*, **6**, 187–197.

Keller, L.F., Grant, P.R., Grant, R. & Petren, K. (2002) Environmental conditions affect the magnitued of inbreeding depression in survival of Darwin's finches. *Evolution*, **56**, 1229–1239.

Koper, N. & Manseau, M. (2009) Generalized estimating equations and generalized linear mixed-effects models for modelling resource selection. *Journal of Applied Ecology*, **46**, 590–599.

Lee, L.F. (1982) Specification error in multinomial logit models: Analysis of the omitted variable bias. *Journal of Econometrics*, **20**, 197–209.

Lee, Y. & Nelder, J.A. (2001) Hierarchical generalized linear models: A synthesis of generalised linear models, random-effects models and structured dispersions. *Biometrika*, **88**, 987–1006.

Lee, Y. & Nelder, J.A. (2004) Conditional and marginal models: another view. *Statistical Science*, **19**, 219–238.

Liang, K.Y. & Zeger, S.L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13–22.

Lindsey, J.K. (1999) *Models for Repeated Measurements*. The Clarendon Press, Oxford University Press, New York, 2 edition.

Lindsey, J.K. & Lambert, P. (1998) On the appropriateness of marginal models for repeated measurements in clinical trials. *Statistics in Medicine*, **17**, 447–469.

Molenberghs, G. & Verbeke, G. (2005) *Models for Discrete Longitudinal Data.* Springer, New York.

Nakagawa, S. & Schielzeth, H. (2013) A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, **4**, 133–142.

Neuhaus, J., Hauck, W. & Kalbfleisch, J. (1992) The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika*, **79**, 755–762.

Neuhaus, J., Kalbfleisch, J. & Hauck, W. (1991) A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *International Statistical Review*, **59**, 25–35.

Overall, J.E. & Tonidandel, S. (2004) Robustness of generalized estimating equation (GEE) tests of significance against misspecification of the error structure model. *Biometrical Journal*, **26**, 203–213.

Pan, W. (2001) Akaike's information criterion in generalized estimating equations. *Biometrics*, **57**, 120–125.

Paradis, E. & Claude, J. (2002) Analysis of comparative data using generalized estimating equations. *Journal of Theoretical Biology*, **218**, 175–185.

Pauler, D.K. (1998) The Schwarz criterion and related methods for normal linear models. *Biometrika*, **85**, 13–27.

Pepe, M.S. & Anderson, G.L. (1994) A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics – Simulation and Computation*, **23**, 939–951.

Pinheiro, J.C. & Bates, D.M. (2000) *Mixed-Effects Models in S and S-Plus.* Springer, New York.

Pinheiro, J.C. & Chao, E.C. (2006) Efficient Laplacian and adaptive Gaussian quadrature algorithms for multilevel generalized linear mixed moels. *Journal of Computational and Graphical Statistics*, **15**, 58–81.

Poncet, B.N., Herrmann, D., Gugerli, F., Taberlet, P., Holderegger, R., Gielly, L., Rioux, D., Thuiller, W., Aubert, S. & Manel, S. (2010) Tracking genes of ecological relevance using a genome scan in two independent regional population samples of Arabis alpina. *Molecular Ecology*, **19**, 2896–2907.

R Core Team (2013) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raudenbush, S.W. (2008) Targets of inference in hierarchical models for longitudinal data. G. Fitzmaurice, M. Davidian, G. Verbeke & G. Molenberghs, eds., *Longitudinal Data Analysis*. Chapman & Hall/CRC, Boca Raton.

Raudenbush, S.W., Yang, M.L. & Yosefc, M. (2000) Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate Laplace approximation. *Journal of Computational and Graphical Statistics*, **9**, 141–157.

Reeves, G.K., Cox, D.R., Darby, S.C. & Whitley, E. (1998) Some aspects of measurement error in explanatory variables for continuous and binary regression models. *Statistics in Medicine*, **17**, 2157–2177.

Ritz, J. & Spiegelman, D. (2004) Equivalence of conditional and marginal regression models for clustered and longitudinal data. *Statistical Methods in Medical Research*, **13**, 309–323.

Rudemo, M., Ruppert, D. & Streibig, J.C. (1989) Random-effect models in nonlinear regression with applications to bioassay. *Biometrics*, **45**, 349–362.

Rue, H., Martino, S. & Chopin, N. (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society Series B (Methodological)*, **71**, 319–392.

Schildcrout, J.S. & Heagerty, P.J. (2005) Regression analysis of longitudinal binary data with time-dependent environmental covariates: bias and efficiency. *Biostatistics*, **6**, 633–652.

Schildcrout, J.S. & Heagerty, P.J. (2007) Marginalized models for moderate to long series of longitudinal binary response data. *Biometrics*, **63**, 322–331.

Simpson, E.H. (1951) The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society Series B (Methodological)*, **13**, 238–241.

Stefanski, L. (1985) The effects of measurement error on parameter estimation. *Biometrika*, **72**, 583–592.

Subramanian, S.V. & O'Malley, A.J. (2010) Modeling neighborhood effects: The futility of comparing mixed and marginal approaches. *Epidemiology*, **21**, 475–481.

Tosteson, T.D., Stefanski, L.A. & Schafer, D.W. (1989) A measurement error model for binary and ordinal regression. *Statistics in Medicine*, **8**, 1139–1147.

Wakefield, J. (2013) *Bayesian and Frequentist Regression Methods.* Springer, New York.

Wedderburn, R.W.M. (1974) Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, **61**, 439–447.

Young, M.L., Preissner, J.S., Quaqish, B.F. & Wolfson, M. (2007) Comparison of subject-specific and population averaged models for count data from cluster-unit intervention trials. *Statistical Methods in Medical Research*, **16**, 167–184.

Zeger, S. & Liang, K.Y. (1986) Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, **42**, 121–130.

Zeger, S., Liang, K.Y. & Albert, P. (1988) Models for longitudinal data: A generalized estimating equation approach. *Biometrics*, **44**, 1049–1060.

Zicus, M.C., Fieberg, J. & Rave, D.P. (2003) Does mallard clutch size vary with landscape composition: a different view. *Wilson Bulletin*, **115**, 409–413.

Zicus, M.C., Rave, D.P., Das, A., Riggs, M.R. & Buitenwerf, M.L. (2006) Influence of land use on mallard nest-structure occupancy. *Journal of Wildlife Management*, **70**, 1325–1333.

Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A. & Smith, G.M. (2009) *Mixed Effects Models and Extensions in Ecology with R.* Springer, New York.

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1: R code to reproduce the mallard nest structures example.

	GLMM1	GEE1	GLMM2	GEE2	λ_1	λ_2
`year1998`	1.70	1.46	1.71	1.42	0.86	0.83
	(0.54)	(0.51)	(0.54)	(0.49)		
`year1999`	1.45	1.25	1.45	1.21	0.86	0.83
	(0.56)	(0.43)	(0.56)	(0.41)		
`period2`	1.33	1.15	1.34	1.10	0.86	0.82
	(0.56)	(0.51)	(0.56)	(0.48)		
`period3`	-1.19	-1.04	-1.01	-0.87	0.87	0.86
	(0.89)	(0.96)	(0.90)	(0.92)		
`period4`	-3.13	-2.72	-2.93	-2.45	0.87	0.84
	(1.23)	(1.35)	(1.26)	(1.29)		
`year1998:period2`	-3.11	-2.70	-3.10	-2.60	0.87	0.84
	(0.81)	(0.67)	(0.81)	(0.63)		
`year1998:period3`	-3.11	-2.68	-3.09	-2.57	0.86	0.83
	(0.88)	(0.83)	(0.88)	(0.78)		
`year1998:period4`	-2.63	-2.29	-2.61	-2.19	0.87	0.84
	(0.89)	(0.85)	(0.89)	(0.80)		
`year1999:period2`	-2.21	-1.92	-2.15	-1.81	0.87	0.84
	(0.79)	(0.56)	(0.79)	(0.52)		
`year1999:period3`	-2.81	-2.43	-2.76	-2.30	0.86	0.83
	(0.89)	(0.69)	(0.89)	(0.64)		
`year1999:period4`	-1.83	-1.58	-1.81	-1.50	0.86	0.83
	(0.85)	(0.77)	(0.85)	(0.73)		
`log(VOM)`	1.58	1.36	1.48	1.23	0.86	0.83
	(0.50)	(0.49)	(0.52)	(0.46)		
`log(size)`	0.68	0.54	–	–	0.80	–
	(0.21)	(0.16)				
$\hat{\sigma}_\tau^2$	1.16	–	1.50	–		
$\hat{\rho}$	–	0.118	–	0.124		

Table 1: Parameter estimates and standard errors (in brackets) of the mallard example. The parameter λ_1 denotes the attenuation factors between the GEE and the GLMM estimates of model 1, while λ_2 gives the respective factors for model 2.

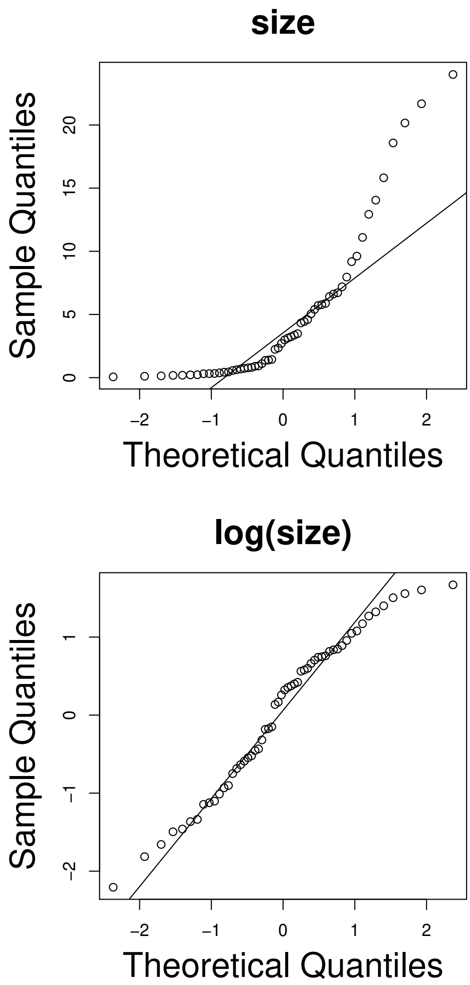

Figure 1: Distribution of the `size` covariate from the mallard data set before (top) and after log transformation (bottom).